Languages adapt to minute differences in their speakers' ecology

Freek Van de Velde

University of Leuven

Quantitative Lexicology and Variational Linguistics – QLVL

Acknowledgements:

Isabeau De Smet

Dirk Pijpops

Katrien Beuls

Main ideas of this talk

- Language change is a function of population structure (historical demography)
- This can be most clearly established by comparing different language families, and on the synchronic level
- But working with proxies and with simulations, we can make educated guesses about the past, within language families

The structuralist era

- Ferdinand De Saussure: 'internal' vs. 'external' linguistics

- Many linguists are wary of digging into external explanations for language change (Lass, Kuryłowicz, Ohala ... see e.g. Woods 2001: 974-975)

  "In view of the confusion and controversies surrounding causes of language change, it is not surprising that some reputable linguists have regarded the whole field as a disaster area, and opted out altogether." (Aitchison 1991: 106)

  "Although language-internal structural correlations such as implicational universals have long been treated as an interesting problem in linguistics, the same cannot be said of attempts to relate facts about language structure to facts about speakers and their environment – variables such as group size, geographical location, genetic makeup, and cultural expectations. (...). In general, (...) attempts to link language structure with extralinguistic factors are almost intrinsically suspect." (Ladd et al. 2015: 227)

Reasons for the wariness

- If morphosyntax is too responsive to demographic change, this potentially undermines universalism, which reeks of crypto-racism and discredited romantic ideas (Herder, Humboldt, Schiller ...) about the deep connection between language and people

- There is no shortage of crackpot theories
  - German tribes using fricatives instead of stops because of their impetous nature
  - German tribes using fricatives instead of stops because of the huffing and puffing in mountainous terrain
  - recently en vogue again: dry climate doesn't sustain tonal languages; having a future tense distinct from the present make you save less money

- General aversion to 'nomothetic' (as opposed to 'idiographic') approaches in linguistics (Roberts & Winters 2012)

Reasons for the wariness (continued)

- Results of research into the impact of demographic factors on language change are unclear:
  - Nettle (1999): smaller languages, faster change
  - Wichmann & Holman (2009): 1° No clear effects: "The test shows mainly negligible effects of population", 2° "... the exception being an apparently faster rate of change in the larger of two closely related variants."
  - Bromham et al. (2015): Higher lexical gain and lower lexical loss in large languages
  - Trudgill (2002: 725): Linguistic change tends to be relatively rapid in high-contact language communities where contact is short-term and/or involves imperfect language learning by adults.

Restoration

- Labovian sociolinguistics
- Contact linguistics
- Evolutionary linguistics

Restoration

- Labovian sociolinguistics
- Contact linguistics

> "In a sense, most of what historical linguists study under the designation "language change" is due to contact." (Thomason 2003: 687)

> "[C]ontact has emerged in recent studies as a more essential element in triggering linguistic innovation than had previously been assumed."
>
> (Drinka 2010: 342)

- However: a lot of studies deal with **borrowing** (also Harris & Campbell 1995; Lucas 2015)
- Wat remains more controversial is **(de)complexification and analyticisation**

> "Language contact, especially when extensive L2 learning is involved, is a main source of complexity reduction (grammar simplification). By definition, such processes involve language change. But complexity reduction is actually at the heart of many type of language change, especially in morphology and syntax."
>
> (Karlsson et al. 2008: viii)

Evolutionary linguistics

- Evolutionary linguistics, as opposed to biolinguistics:
  - Languages are considered as complex adaptive systems (Holland 1992; Briscoe 2000; Bybee 2010; Steels 2011a), with emergent properties (Keller 1994)
  - Adaptation assumes the existence of selection pressures and fitness
  - Language adapt to the niche occupied by the speech community (Kusters 2003 Christiansen & Chater 2008; Lupyan & Dale 2010; Dale & Lupyan 2012; Bentz & Christiansen 2013).
  - The difference between speech communities revolves around different learning constraints.
  - Evolutionary linguistics is nomothetically oriented (as opposed to the often more idiographic approaches in mainstream historical linguistics)

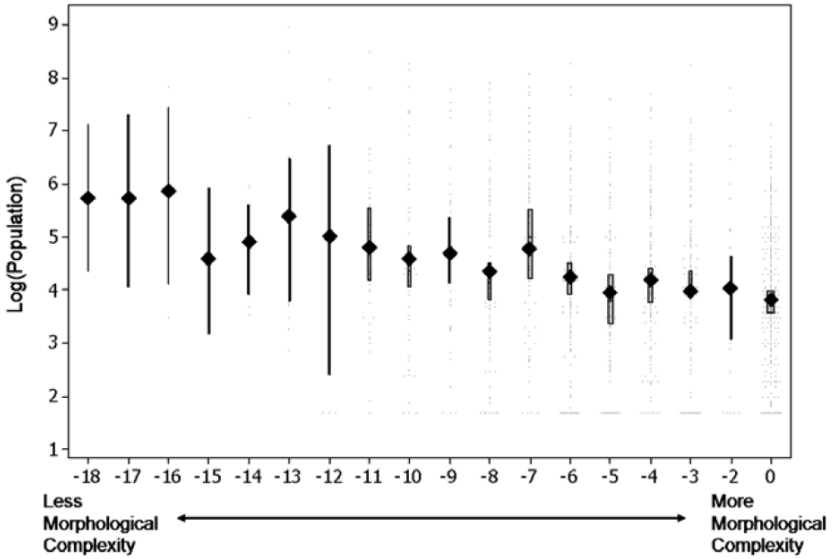
Linguistic niche hypothesis

Lupyan & Dale (2010)

- **Esoteric languages**: morphologic complexity, redundancy, synthetic ~ favouring L1 acquisition, typically smaller languages

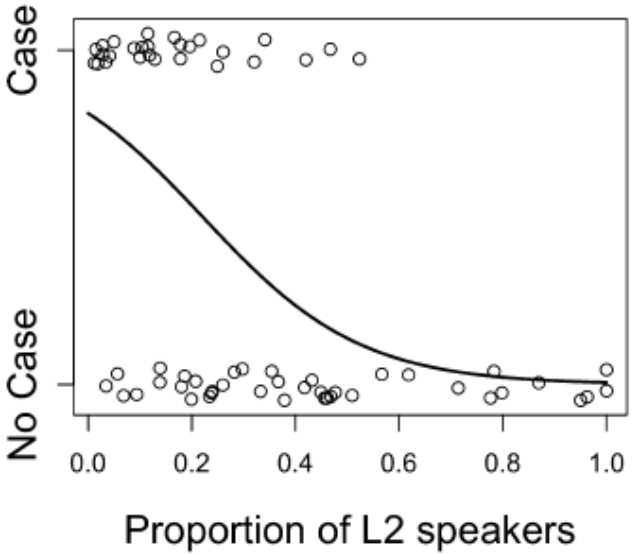
  "children are "little inflection machines"" (Wexler 1998: 43)

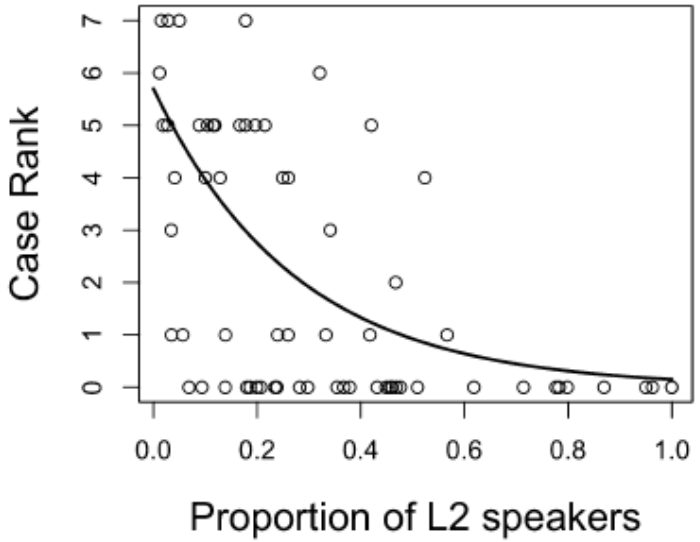
- **Exoteric languages**: analytic-syntactic complexity, transparency, analytic ~ favouring L2 acquisition, typically bigger languages



⇦ Lupyan & Dale 2010

⇩ Bentz & Winter 2013





From synchrony to diachrony

- Change: (de)complexification and analyticisation
- Two big, related themes with a time-honoured scholarly interest:
  1. Morphological types
  2. Demographic correlations

Types of languages

- Morphological types: isolating, agglutinative, fusional, polysynthetic, introflexive ... (disagreement and confusion, see Bickel & Nichols 2013)

- Analytic vs. synthetic:

  "En Europe les langues dérivées du latin, et l'anglais, ont une grammaire tout analytique (...) synthétiques dans leur origine (...) elles penchent fortement vers les formes analytiques" (Von Schlegel 1846: 161, cited in Szmrecsanyi 2012)

- What happened in Europe? Adults hijacked the language.

Demographic explanations for the analytic-synthetic difference

- Lupyan & Dale (2010), Bentz & Winter (2013): synchronic quantitative evidence for Schlegel's diachronic claim

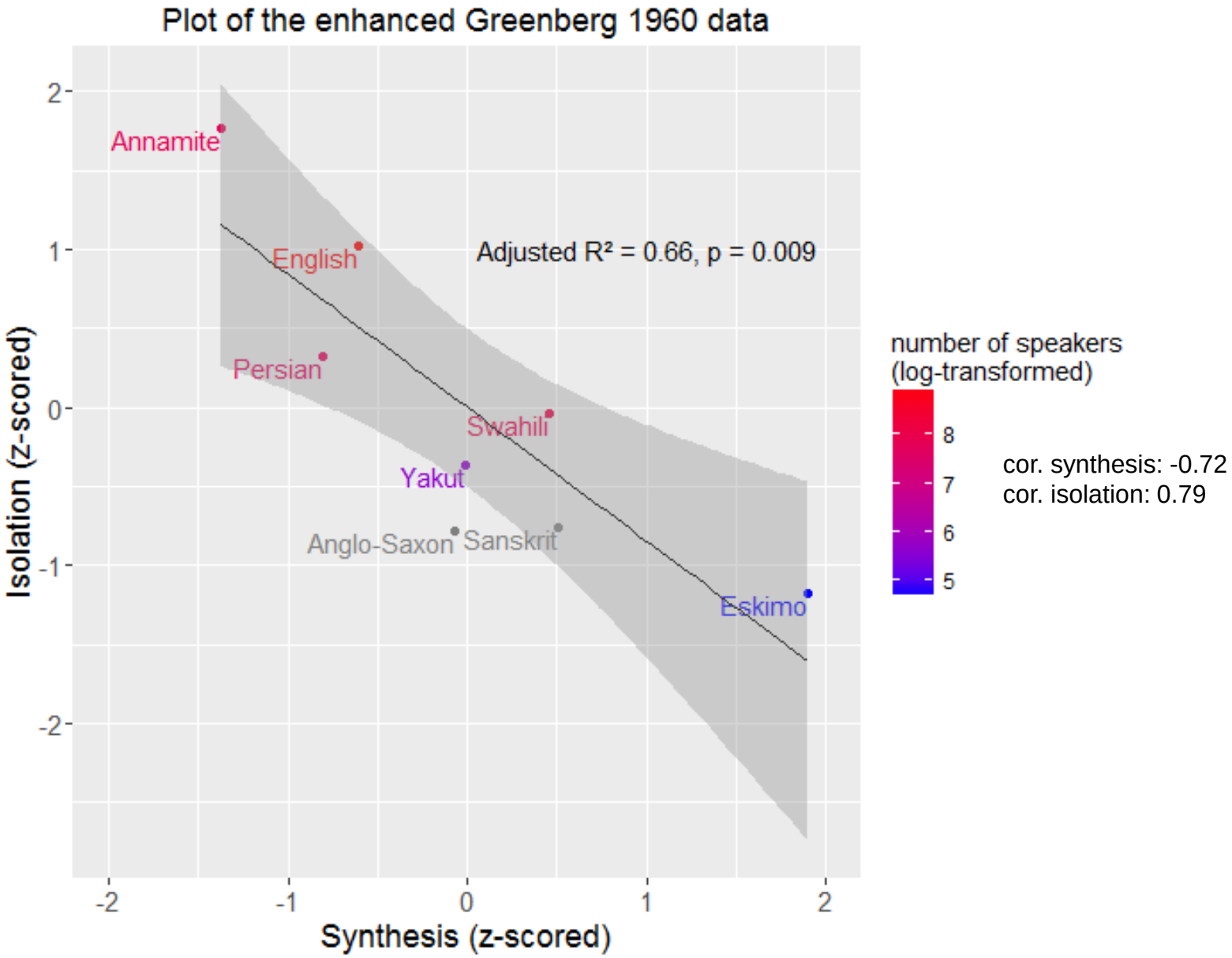
- What about diachronic quantitative evidence? (see Kusters 2003; Szmrecsanyi 2012; Carlier et al. 2012; Haspelmath, forthc.; Haspelmath & Michaelis, forthc.)

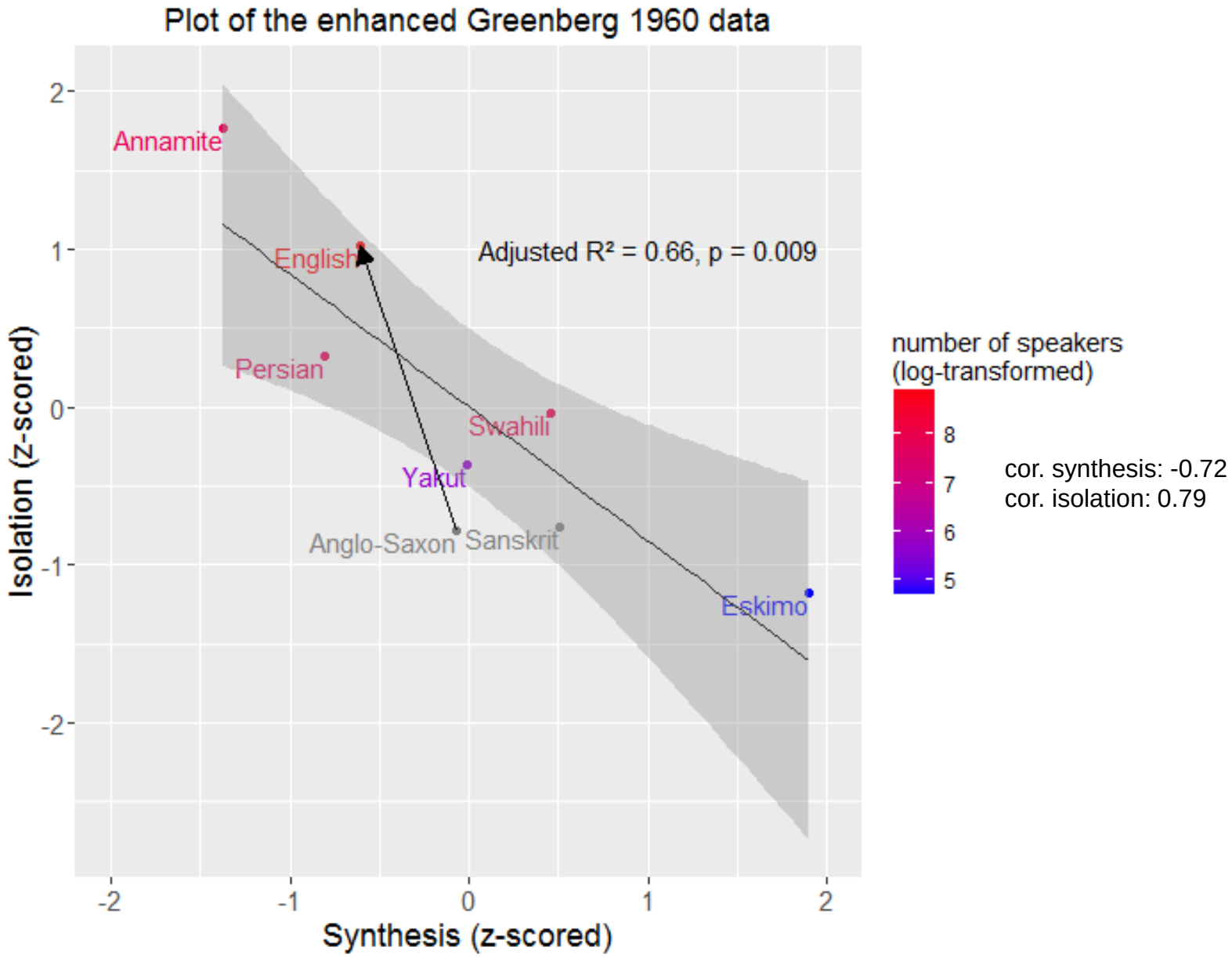
From synchrony to diachrony

Greenberg (1960)

- Index of synthesis (proportion of morphemes to words)
- Index of isolation (proportion of word order as a grammatical marker to the total number of nexus)
- Along with a number of other indices (Index of agglutination, Index of compounding, Index of inflection, Index of prefixation ...).
- Calculated on 100 word stretches of different languages (labour-intensive):

	Sanskrit	Anglo-Saxon	Persian	English	Yakut	Swahili	Annamite	Eskimo
Synthesis	2.59	2.12	1.52	1.68	2.17	2.55	1.06	3.72
Isolation	.16	.15	.52	.75	.29	.40	1.00	.02


Plot of the enhanced Greenberg 1960 data
Annamite
English
Persian
Swahili
Yakut
Anglo-Saxon
Sanskrit
Eskimo
Adjusted R² = 0.66, p = 0.009
Isolation (z-scored)
Synthesis (z-scored)
2
1
0
-1
-2
-2
-1
0
1
2
number of speakers
(log-transformed)
8
7
6
5
cor. synthesis: -0.72
cor. isolation: 0.79


Plot of the enhanced Greenberg 1960 data
Annamite
English
Persian
Swahili
Yakut
Anglo-Saxon
Sanskrit
Eskimo
Adjusted R² = 0.66, p = 0.009
Isolation (Z-scored)
Synthesis (z-scored)
2
1
0
-1
-2
number of speakers
(log-transformed)
8
7
6
5
cor. synthesis: -0.72
cor. isolation: 0.79

From synchrony to diachrony

Haspelmath & Michaelis (forthc.), Haspelmath (forthc.):

- 'word' is a problematic concept
- new operational definition:

> "Analytic pattern: a morphosyntactic pattern that was created from lexical or other concrete material and that is in functional competition with (and tends to replace) an older (synthetic) pattern"
>
> "Thus, the term *analytic* should be understood as roughly meaning "freshly regrammaticalized". This definition works, because all patterns that have traditionally been called "analytic" are known to have been created from lexical or other concrete material; there does not seem to be any other way in which such patterns can come about. This definition is somewhat broader than the traditional purely synchronic definition, in that it also includes cases like the English past-tense marker *-ed* as in *play-ed*, which is generally thought to be a much newer pattern than the old pattern represented by ablauting verbs such as *sing/sang, write/wrote* (...) because it is based on a refunctionalization."

From synchrony to diachrony

Haspelmath & Michaelis (forthc.), Haspelmath (forthc.):

- 'word' is a problematic concept
- new operational definition:

> "Analytic pattern: a morphosyntactic pattern that was created from lexical or other concrete material and that is in functional competition with (and tends to replace) an older (synthetic) pattern"
>
> "Thus, the term *analytic* should be understood as roughly meaning "freshly regrammaticalized". This definition works, because all patterns that have traditionally been called "analytic" are known to have been created from lexical or other concrete material; there does not seem to be any other way in which such patterns can come about. This definition is somewhat broader than the traditional purely synchronic definition, in that it also includes cases like the English past-tense marker *-ed* as in *play-ed*, which is generally thought to be a much newer pattern than the old pattern represented by ablauting verbs such as *sing/sang, write/wrote* (...) because it is based on a refunctionalization."

Case study: weak preterites

- Germanic languages have two morphological strategies for building preterites (not counting analytic perfects, *he has written a book*):

1. Strong inflection:
   - English *sing – sang*
   - Ablaut, based on Indo-European aspectual system (perfect > preterite)

PIE root *b^h_id^h-	e-grade (present)	o-grade (perfect)
Greek	*peíth-omai*	*$_2$pé-poith-a*
Gothic	*beid-an*	**baid-* (PGm ă < PIE ŏ)

2. Weak inflection
   - English *work – worked*
   - Dental suffix, based on a analytic formation [VERB + *d^heh_1-, *d^hoh_1- ('did')]

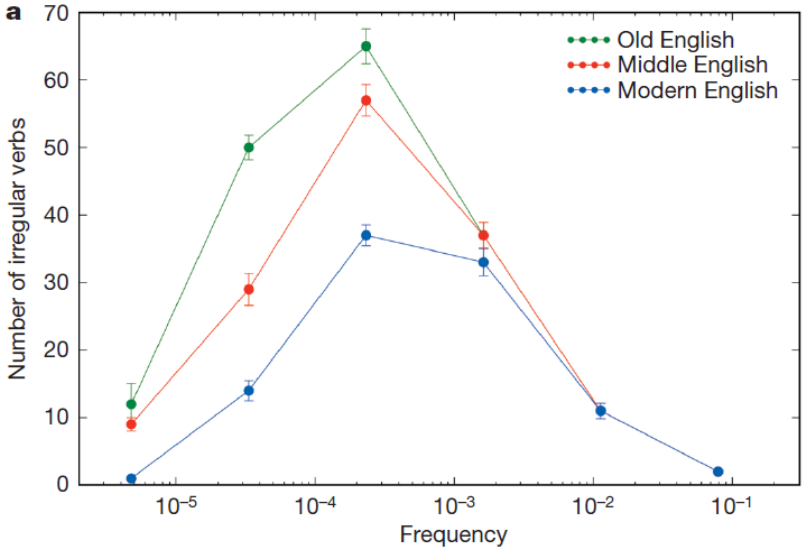
Gothic	*beid-an*	**baid-*
Dutch	*beid-en (~ †bijden)*	*beid-de*

Case study: weak preterites

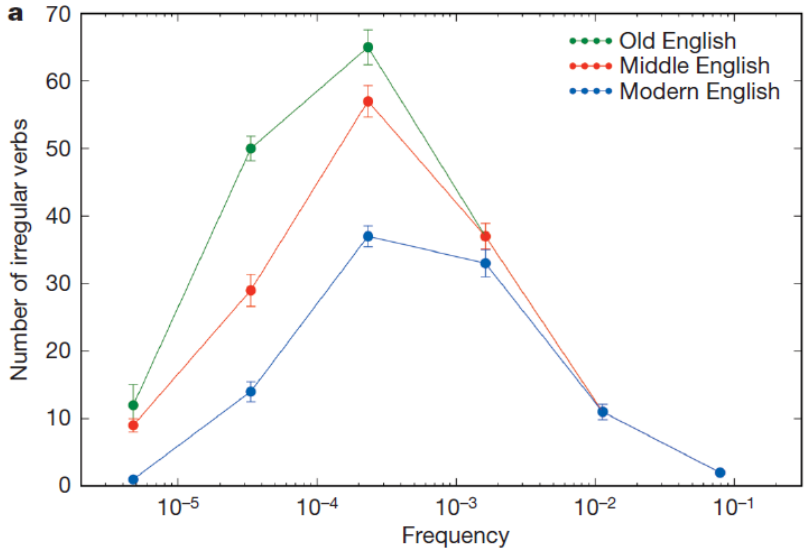
- Various changes occur:
  - irregularisation (Eng. *buy* – *bought*)
  - one strong ablaut class to another (Du. *heffen* – *hief* < *hoef* (Germ. *hob*, *hub*))
  - weak to strong (Du. *vragen* – *vroeg* (vs. Germ. *fragte*))
  - strong to weak (Eng. *carve* – *carved* < *cearf* (Du. *kerfde* < *karf*))

⇨ Long-term drift, over many centuries

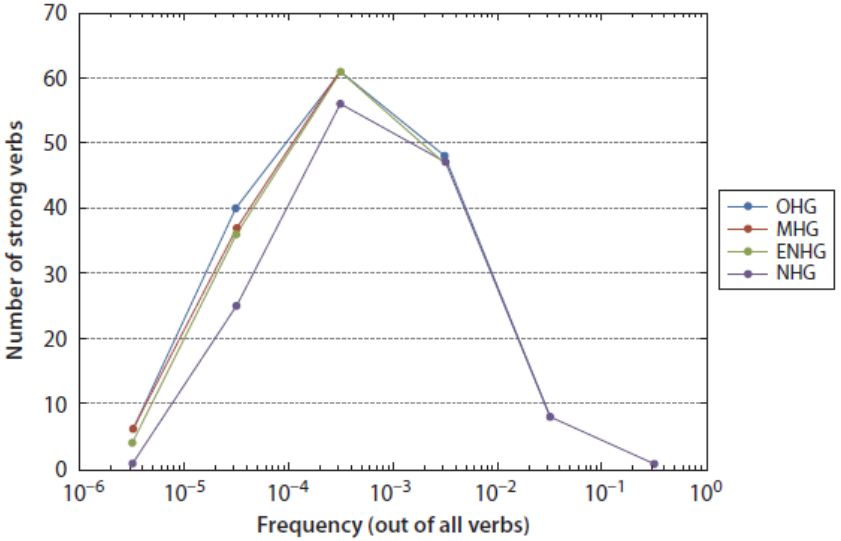
ENGLISH: Lieberman et al. 2007



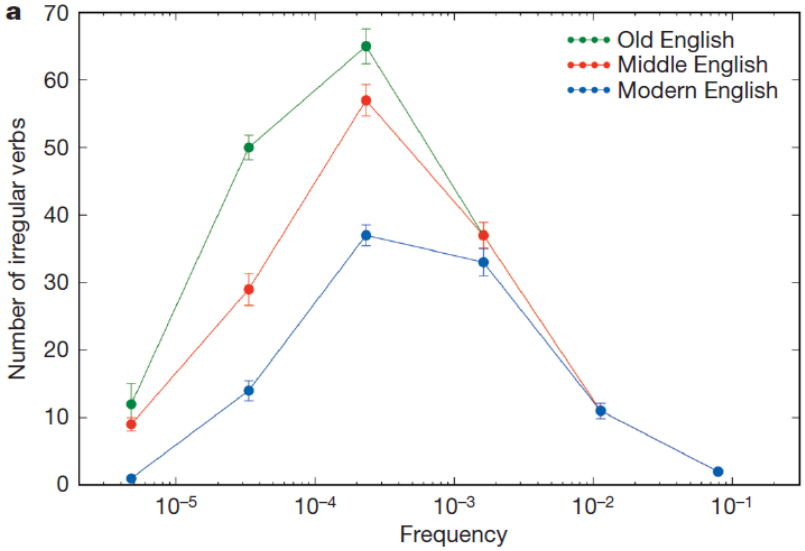
ENGLISH: Lieberman et al. 2007



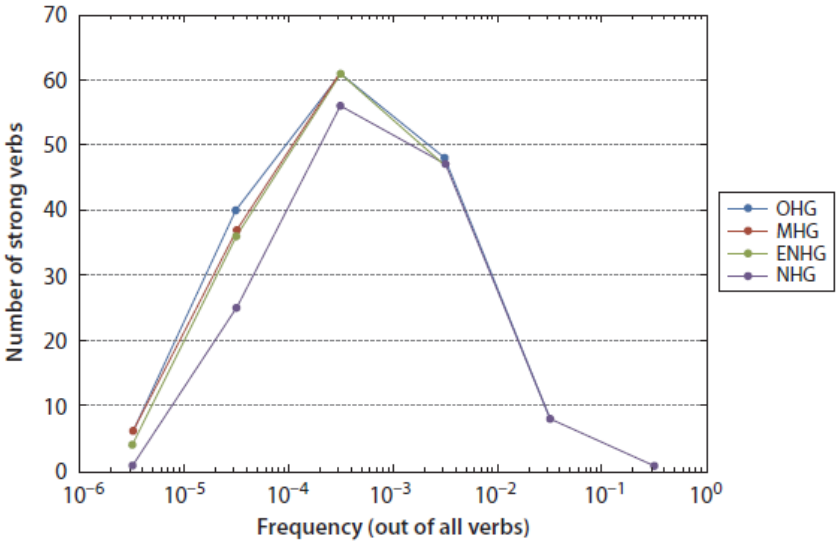
GERMAN: Carroll et al. 2012



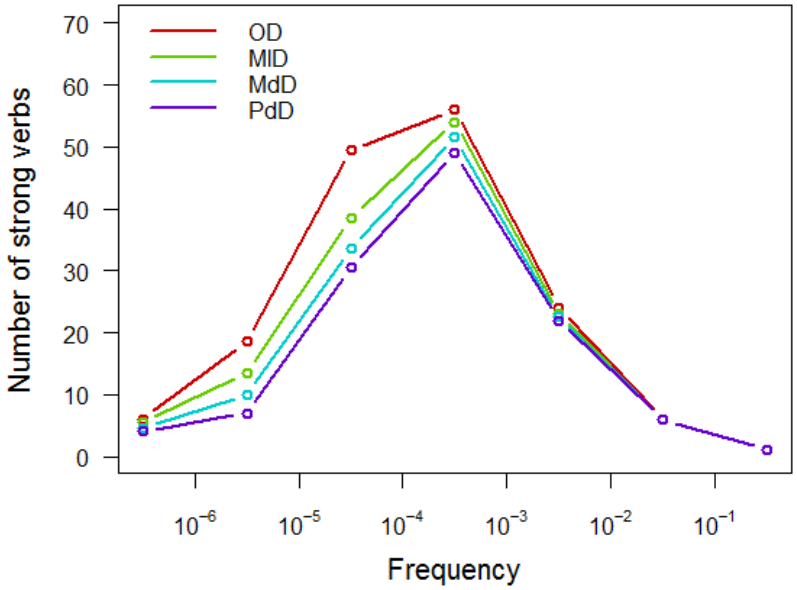
ENGLISH: Lieberman et al. 2007



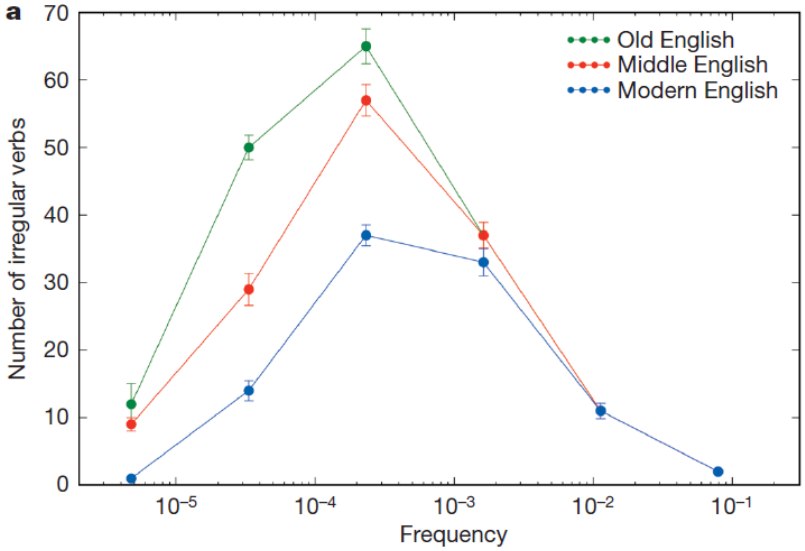
GERMAN: Carroll et al. 2012



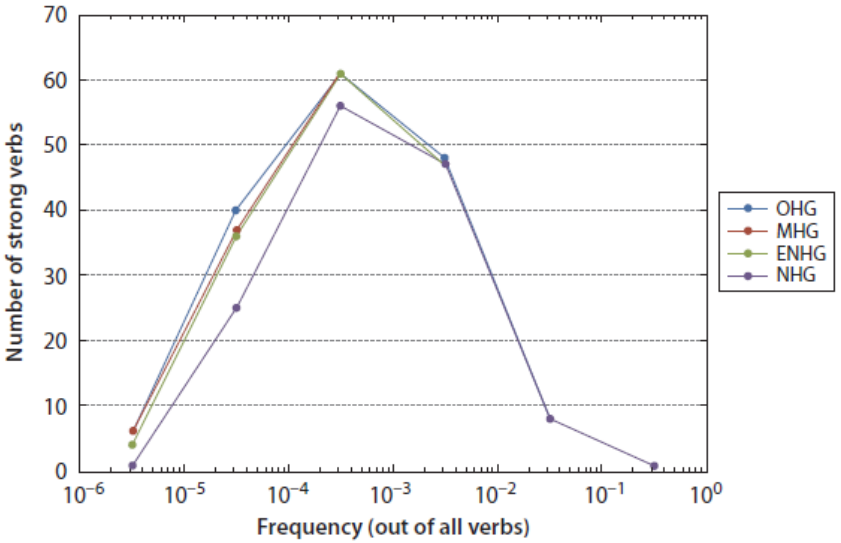
DUTCH



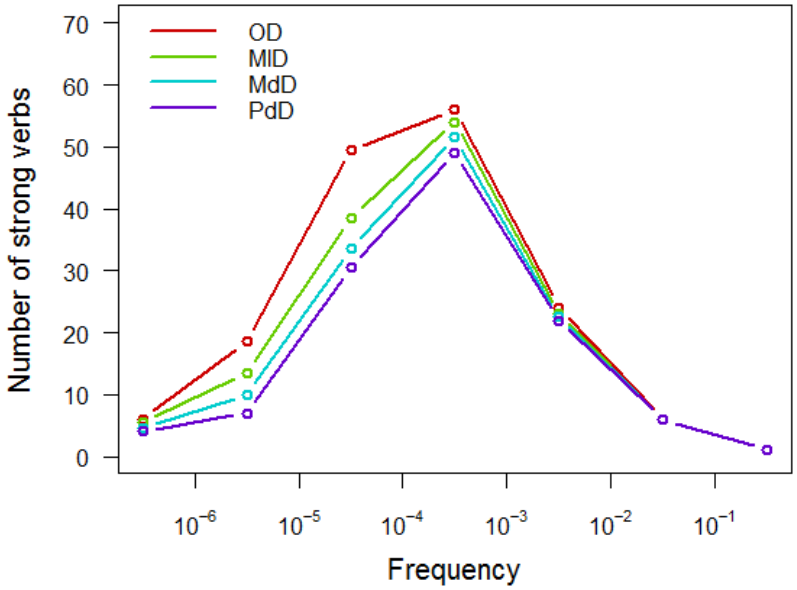
ENGLISH: Lieberman et al. 2007

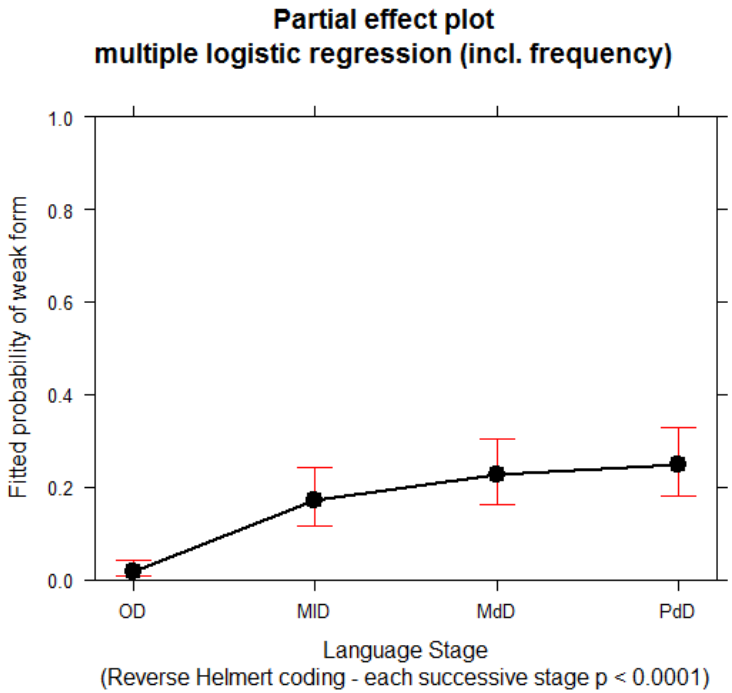


GERMAN: Carroll et al. 2012

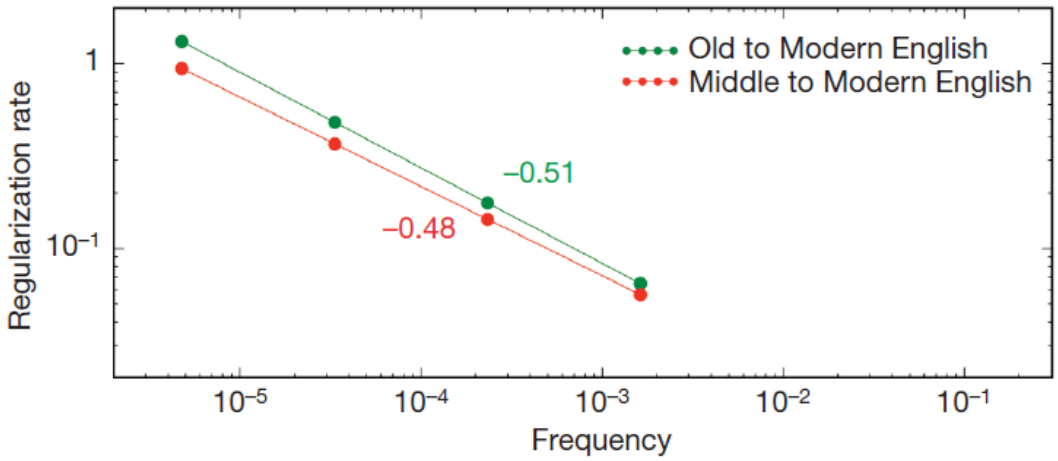


DUTCH



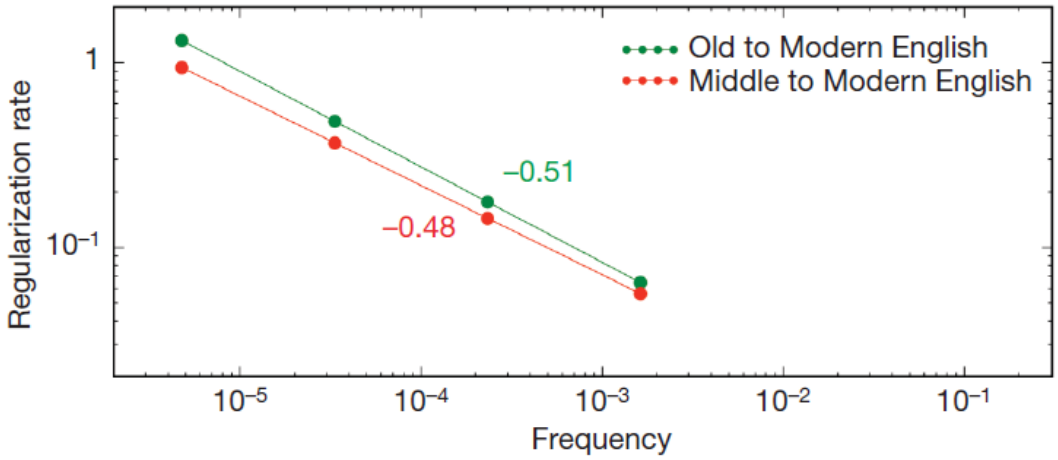


Lieberman et al. 2007: Constant rate of regularisation through time, only dependent on frequency



⇨ lines follow the same power law curve (linear on log-log plot) and overlap

Lieberman et al. 2007: Constant rate of regularisation through time, only dependent on frequency

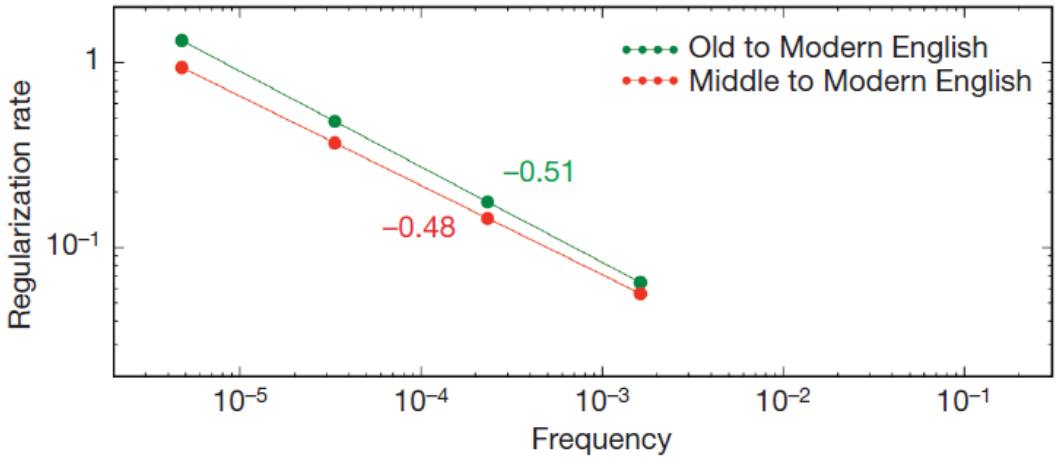


⇨ lines follow the same power law curve (linear on log-log plot) and overlap

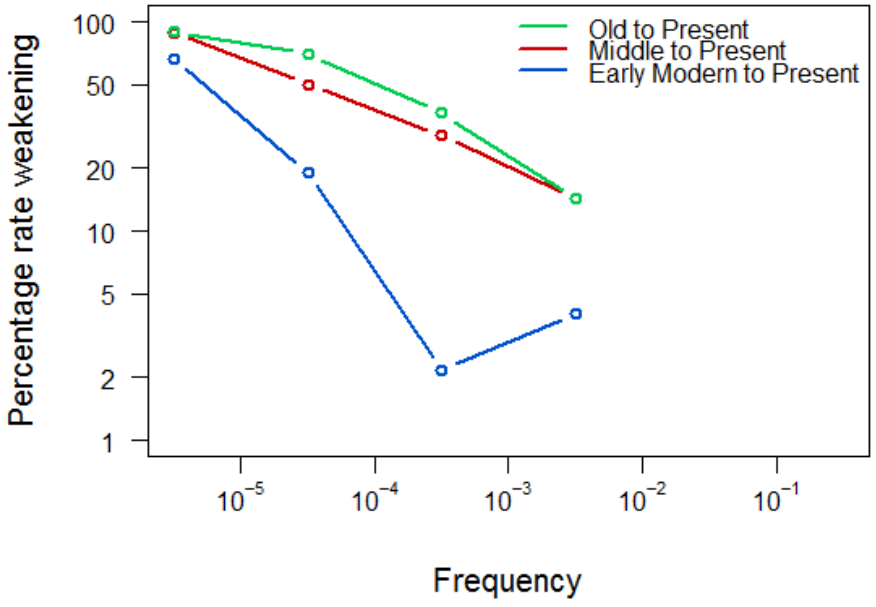
Lieberman et al. 2007: three measurement points:

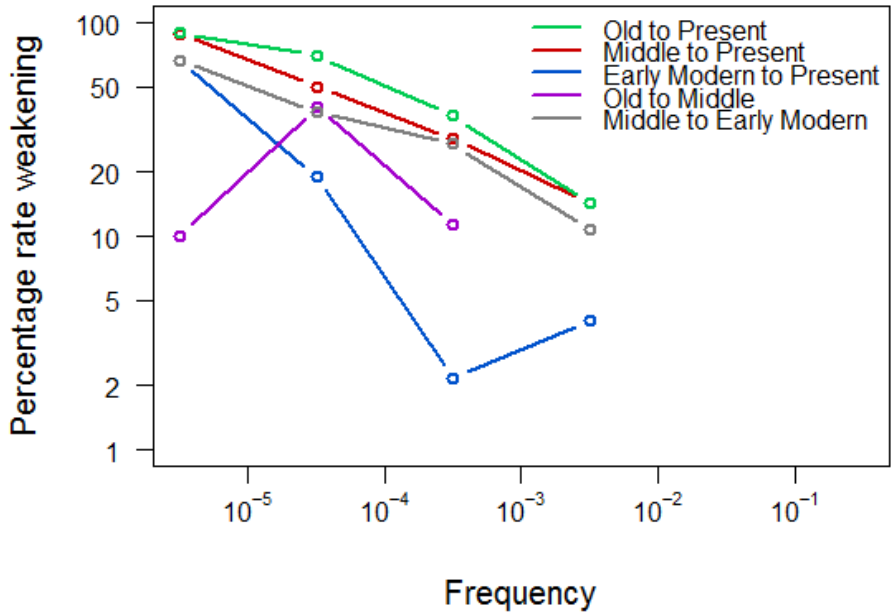
Replication with fourth measurement point:

Lieberman et al. 2007: Constant rate of regularisation through time, only dependent on frequency

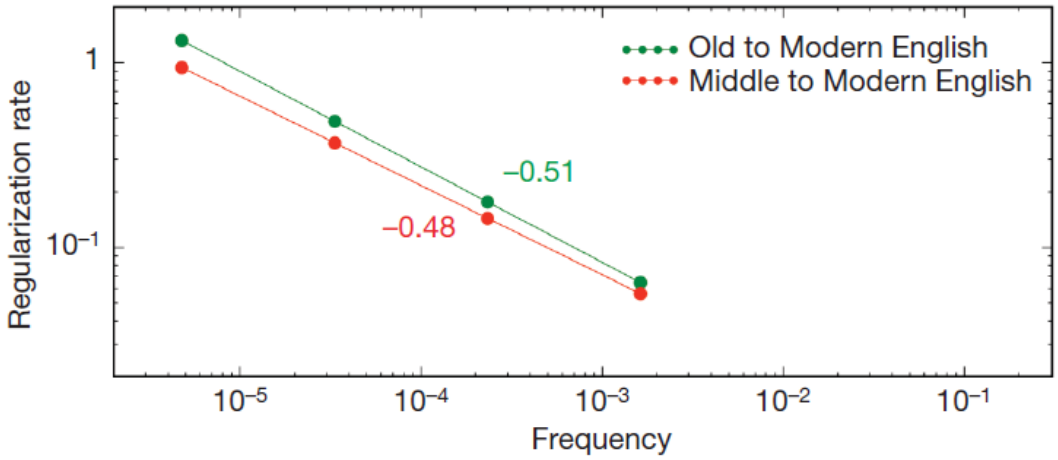


But the constant rate breaks down when we add an extra measurement point for E. Mod. Eng.:

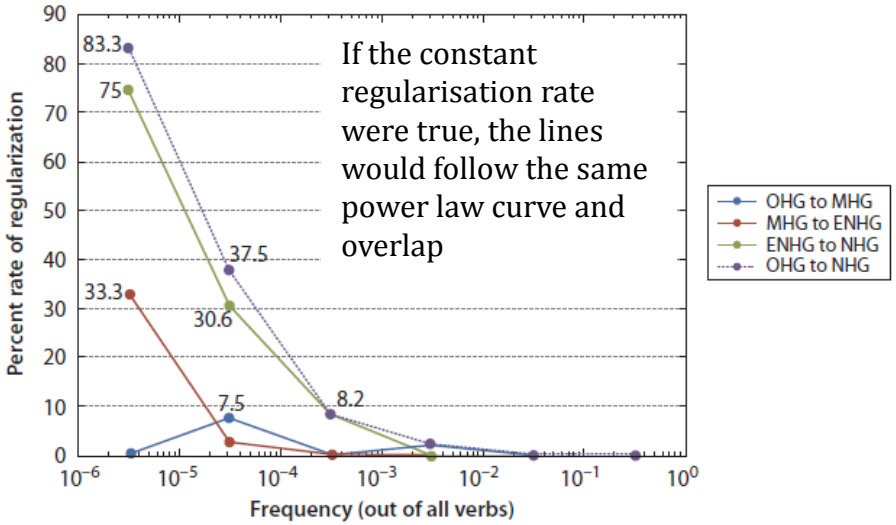




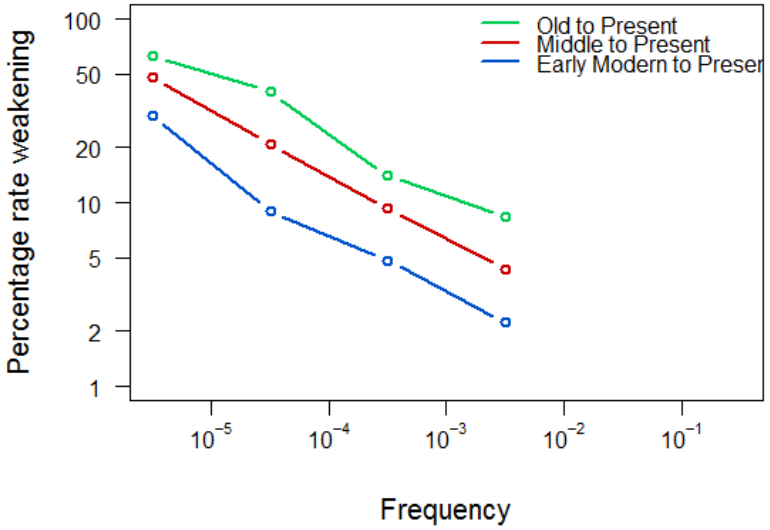
Lieberman et al. 2007: Constant rate of regularisation through time, only dependent on frequency



Carroll et al. 2012: Constant rate does not work for German



... neither for Dutch



Historical demographic data

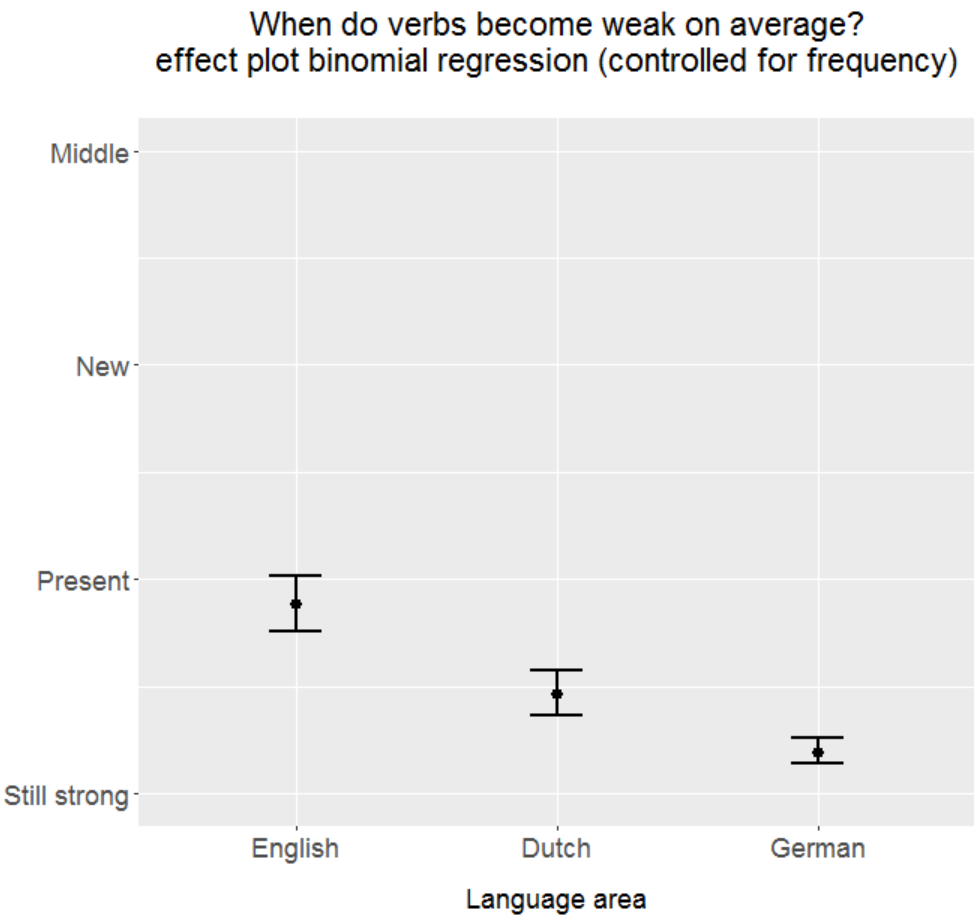
- The change in the preterite does not follow an iron law
- Can the regularisation of the preterite be attributed to demographic factors?

Historical demographic data

- The change in the preterite does not follow an iron law
- Can the analyticisation of the preterite be attributed to language contact?
- Through dialect leveling, koineisation (Kerswill 2002):

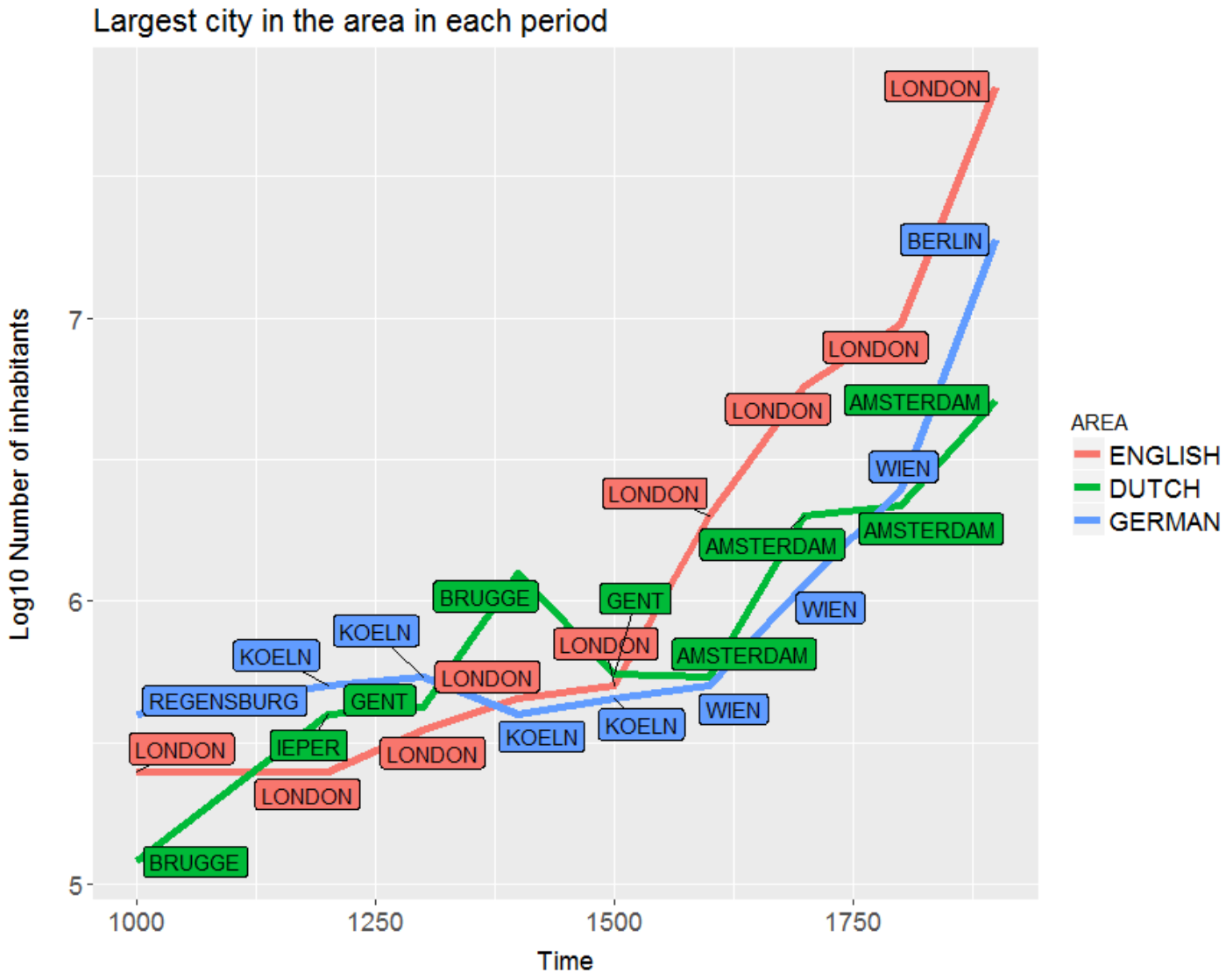
  "[A] social characteristic with structural consequences is dialect or language contact. Increased exposure to different varieties often – though not always – corresponds to patterns of morphological and other leveling or simplification (...). The ENHG period, when verb regularization picks up dramatically in the history of German, is a period notable for increased geographical mobility, in particular urbanization." (Carroll et al. 2012: 169)
- Problem: no direct quantitative data on migration
- Urbanisation is a workable proxy, as:
  - The growth in cities is too large to have come about through natural births, and suggests immigration (Howell 2006: 208)
  - Language change propagates in major urban centres.
  - Language diversity was higher in pre-industrial or industrialising cities than in 20th century cities (E.g. French in France: 1790: 20% (De Certeau et al. 2002), 1893: 75% (Weber 1991: 95))
  - Dialects were often mutually unintelligible

Dutch between English and German (Van Haeringen 1956)



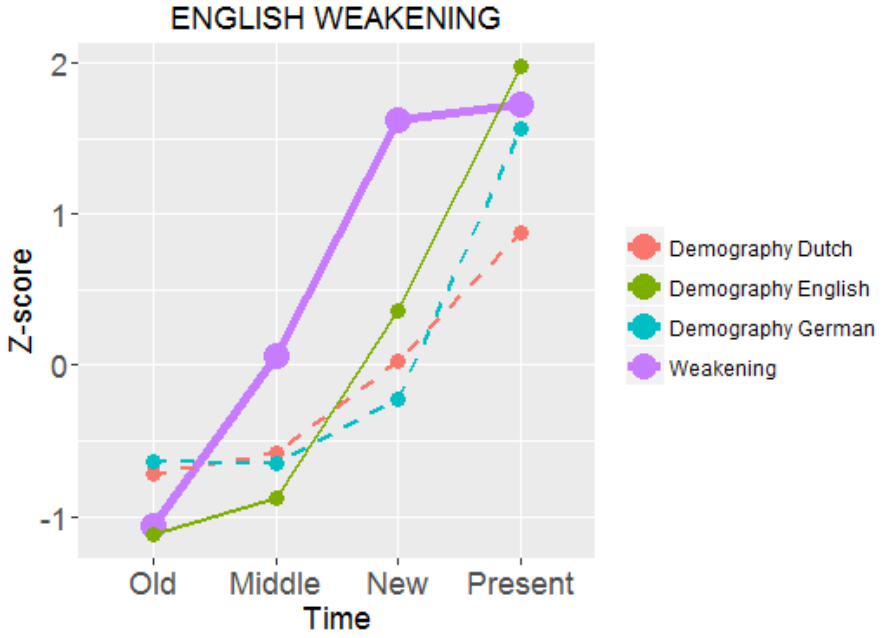
Datasets:

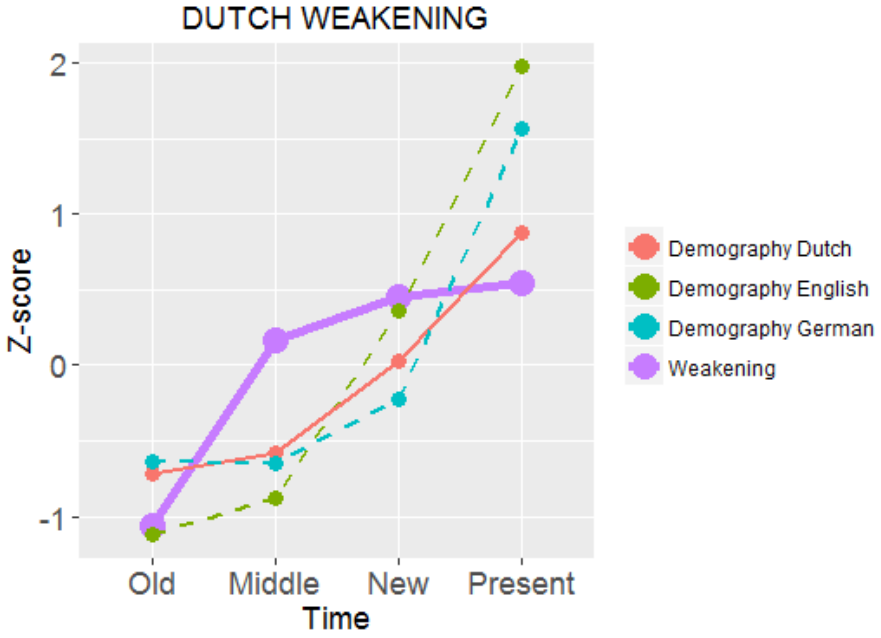
- English: Lieberman et al. (freely available)
- German: Ryan Carroll, Ragnar Svare and Joseph Salmons (shared their data)
- Dutch: Isabeau De Smet

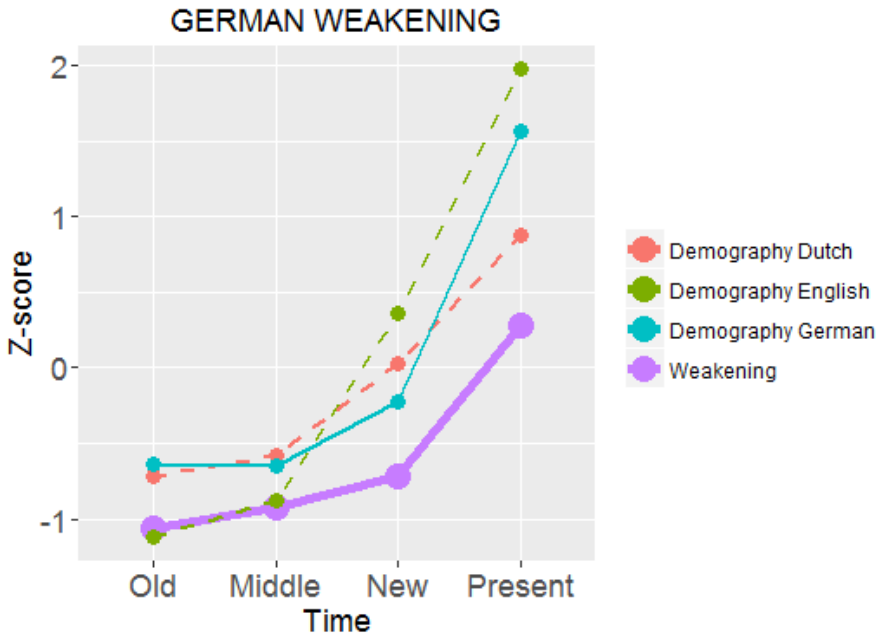

Largest city in the area in each period
Log10 Number of inhabitants
7
6
5
1000
1250
1500
1750
Time
LONDON
BERLIN
LONDON
AMSTERDAM
LONDON
WIEN
LONDON
AMSTERDAM
AMSTERDAM
BRUGGE
GENT
WIEN
KOELN
LONDON
AMSTERDAM
KOELN
LONDON
REGENSBURG
GENT
WIEN
KOELN
KOELN
LONDON
IEPER
LONDON
LONDON
BRUGGE
AREA
ENGLISH
DUTCH
GERMAN

Is this correlation spurious?

- Three arguments why it is not:
  - Closer view on the dynamics of the three systems through time
  - Closer view on the within-area dynamics
  - Agent-based simulation (Pijpops et al. 2015)
- All confirm the role of demography

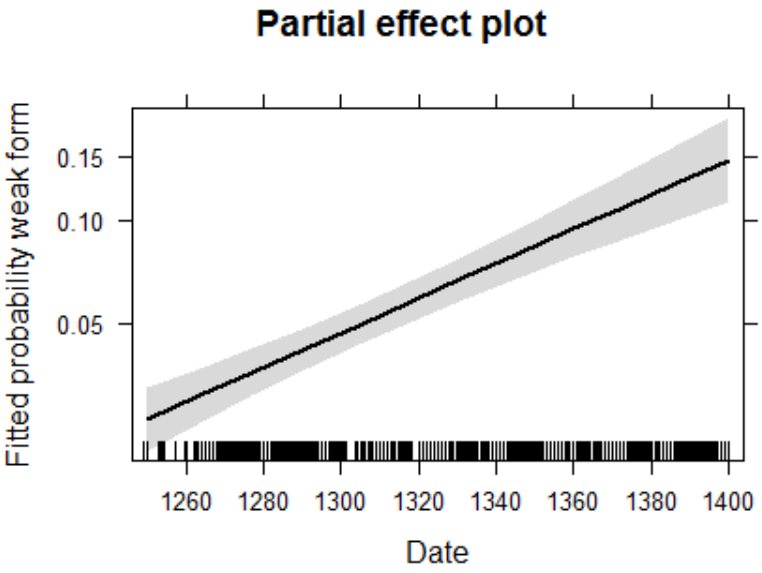


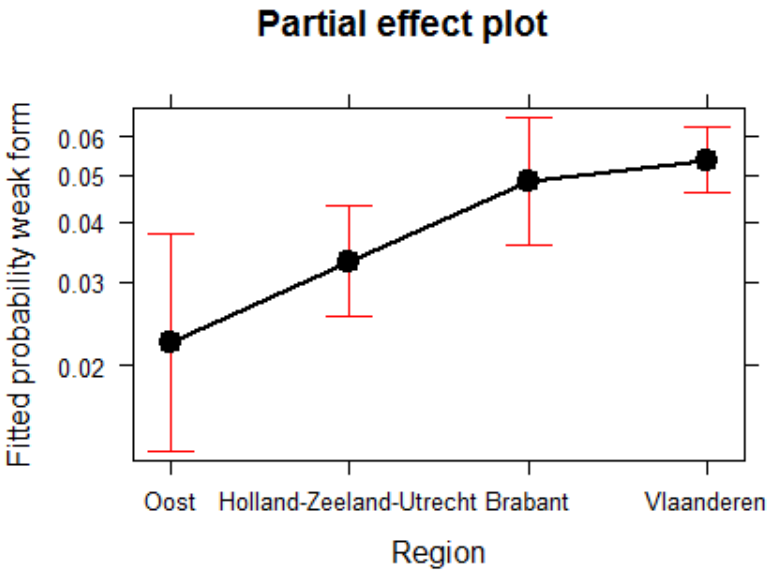


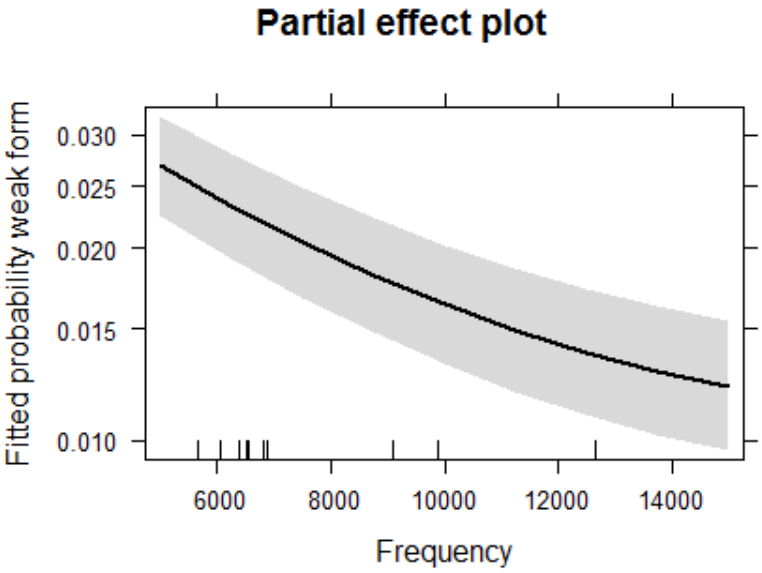


Pearson corr.		Weakening		
		English	Dutch	German
Demography	English	0.87 (p=0.13)	0.72 (p=0.28)	0.96 (p=0.03)
	Dutch	0.86 (p=0.31)	0.72 (p=0.28)	0.97 (p=0.03)
	German	0.69 (p=0.31)	0.56 (p=0.44)	0.99 (p=0.01)

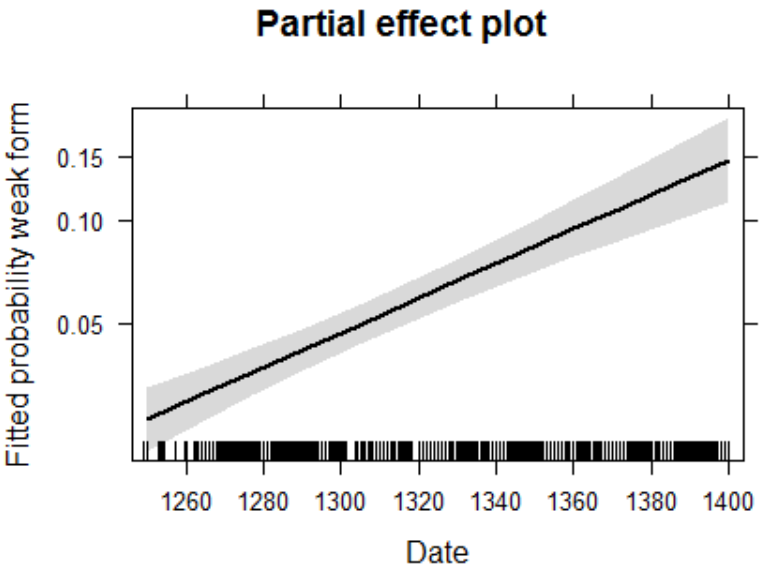
From grammars to corpora: 13th and 14th century Dutch

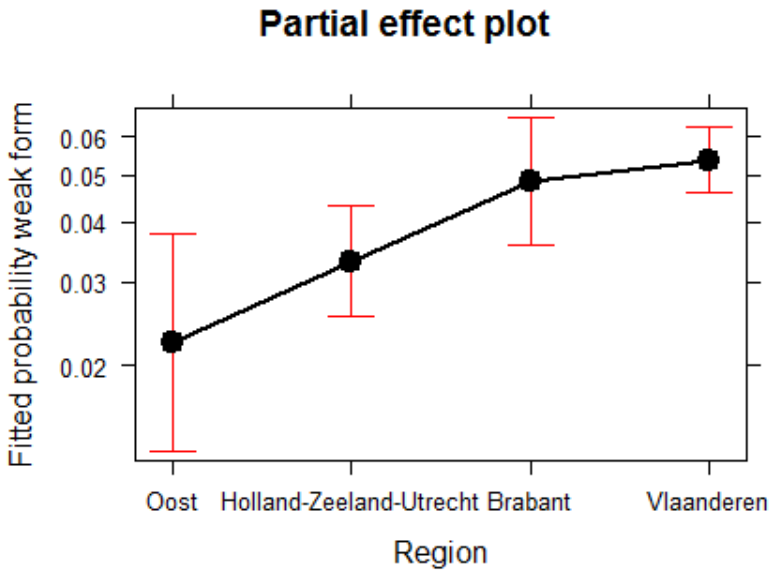


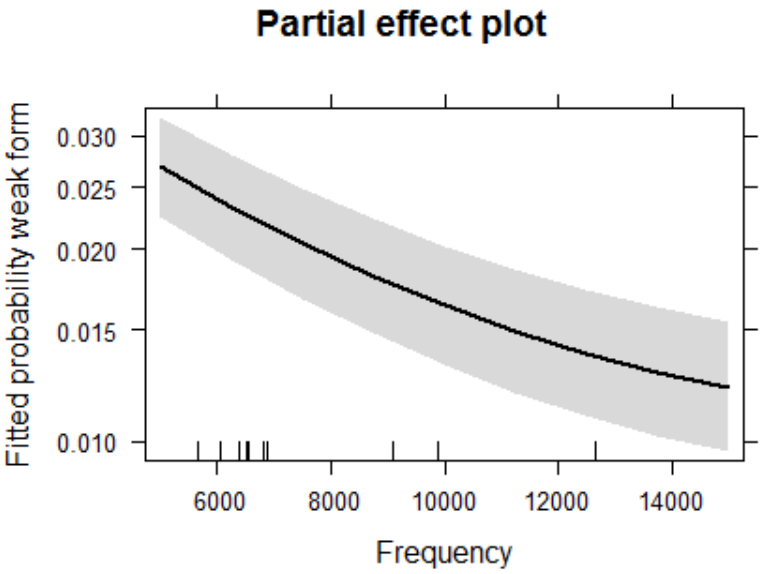


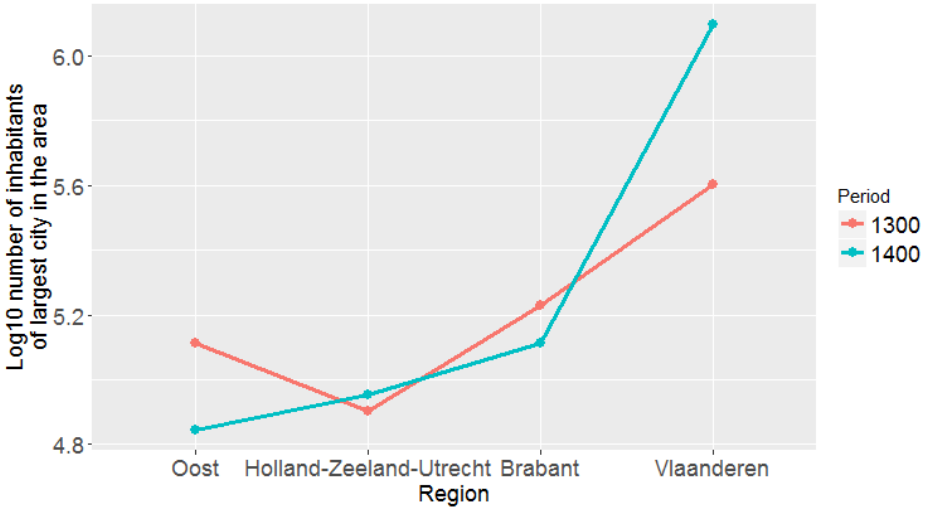


From grammars to corpora: 13th and 14th century Dutch









In silico simulation: agent-based model

(Pijpops, Beuls & Van de Velde 2015)

- Standard method in evolutionary biology, economics etc.
- Recently also applied in linguistics (Dale & Lupyan 2012; Steels 2016 ...)
- Interactions between software agents, equipped with grammar and lexicon
- To see emergence of trends

In silico simulation: agent-based model

- What do we put in?
  - Single, generally applicable weak suffix vs. multiple strong classes
  - Weak suffix has lower type and token frequency to any individual ablaut class (super-conservative)
  - Verbs have a realistic (Zipfian) frequency distribution
  - Agents are gradually replaced (birth & death)
  - Lexical replacement (work in progress)

In silico simulation: agent-based model

Parameters:

- Number of series: 20
- Number of agents: 100
- Time: 5.000.000 times units (average interactions per agent)
- Replacement rate: 1/5.000, 1/10.000, 1/20.000, 1/100.000
- Replacement number: 1
- Verbal replacement: none

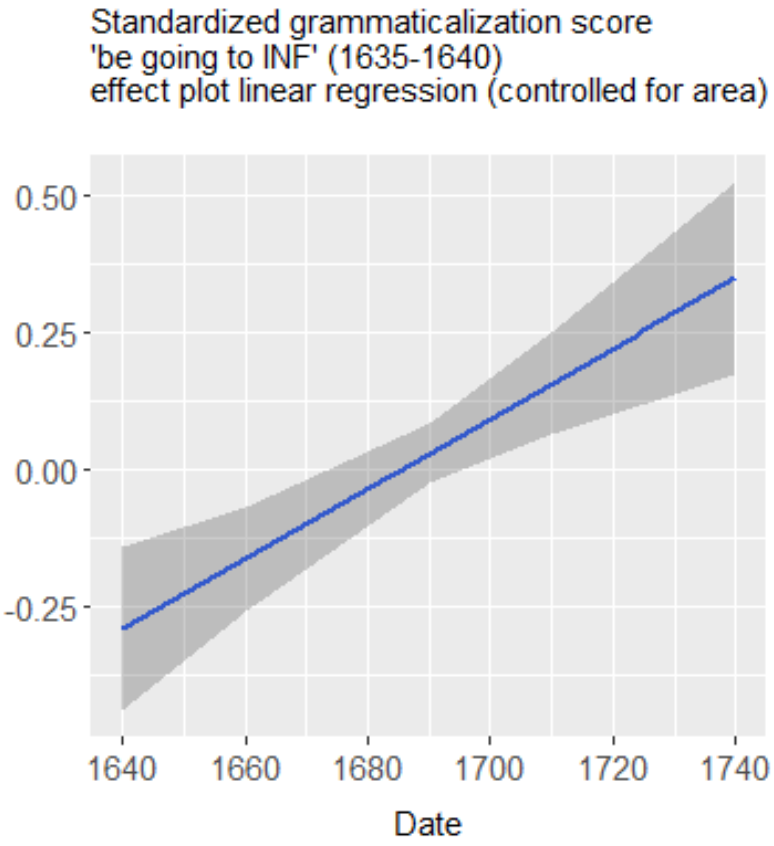
Pijpops, Beuls & Van de Velde (2015)

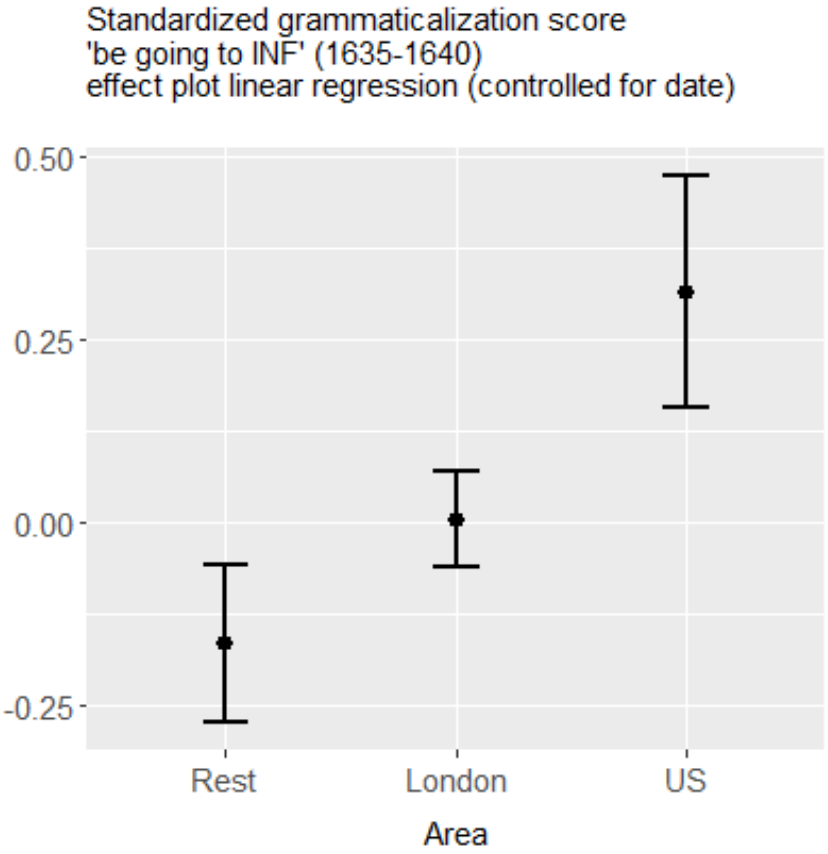
What kind of acquisition?

- Remember that we are (also) dealing with L2 acquisition here:
  - There is no one-to-one vertical transmission from old to new agents, but also horizontal transmission
  - All newly introduced agents have equal speech rights, and have the same influence in accommodation
  - Newly introduced agents are not initially confined to rote learning, but can extract rule-like patterns right away (see Beuls et al. 2010 for a stepwise learning model).
  - Agents have an adult 'world of events' that they want to speak about. They do not start out with simpler verbs like 'drink'. They follow the adult Zipfian distribution of verbs.
- Note that we cannot allow newly introduced agents to fall back on the analytic strategy (weak inflection), as that would give an unduly advantage to the weak inflection

Language contact

- Joint work with Peter Petré (U. of Antwerp)


Standardized grammaticalization score
'be going to INF' (1635-1640)
effect plot linear regression (controlled for area)
0.50
0.25
0.00
-0.25
1640
1660
1680
1700
1720
1740
Date


Standardized grammaticalization score
'be going to INF' (1635-1640)
effect plot linear regression (controlled for date)
0.50
0.25
0.00
-0.25
Rest
London
US
Area

Conclusions

- Language change is a function of historical demography
- between languages of different families
- between languages of the same family (English-Dutch-German)
- within a language (Dutch, English)

Aitchison, J. 1991. *Language change: progress or decay*. 2nd edn. Cambridge: Cambridge University Press.
Bentz, C. & B. Winter. 2013. Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change* 3: 1-27.
Beuls, K., K. Gerasymova & R. van Trijp. 2010. 'Situated learning through the use of language games: solving the Hungarian agreement puzzle'. In: *Proceedings of the 19th Annual Machine Learning Conference of Belgium and The Netherlands (BeNeLearn)*.
Bickel, B. & J. Nichols. 2013. 'Fusion of Selected Inflectional Formatives'. In: M. Dryer & M. Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
Briscoe, E.J. 2000. 'Evolutionary perspectives on diachronic syntax'. In: S. Pintzuk et al. (eds.), *Diachronic syntax: models and mechanisms*. Oxford: Oxford University Press.
Bromham, L., X. Huaa, T.G. Fitzpatrick & S.J. Greenhill. 'Rate of language evolution is affected by population size'. *PNAS* 112: 2097-2102.
Carlier, A., W. De Mulder & B. Lamiroy. 2012. 'Introduction: the pace of grammaticalization in a typological perspective'. *Folia Linguistica* 46(2): 287-301.
Carroll, R., R. Svare & J. Salmons. 2012. 'Quantifying the evolutionary dynamics of German verbs'. *Journal of Historical Linguistics* 2: 153-172.
Christiansen, M.H. & N. Chater. 2008. 'Language as shaped by the brain'. *Behavioral and Brain Sciences* 31(5): 489-509.
De Certeau, M., D. Julia & J. Revel. 1975. *Une politique de la langue. La révolution française et les patois: l'enquête de Grégoire*. Paris: Gallimard
Drinka, B. 2010. 'Language contact'. In: S. Luraghi & V. Bubenik (eds.), *The Continuum companion to historical linguistics*. London: Continuum. 325-345.
Greenberg, J. 1960. 'A Quantitative Approach to the Morphological Typology of Language'. *International Journal of American Linguistics* 26(3): 178-194.
Harris, A.C. & L. Campbell. 1995. *Historical syntax in cross-linguistic perspective.* Cambridge: Cambridge University Press.
Haspemath, M. Forthcoming. 'Revisiting the anasynthetic spiral'. In: H. Narrog & B. Heine (eds.), *Grammaticalization in a typological perspective.*
Haspelmath, M. & S.M. Michaelis. Forthc. 'Analytic and synthetic: typological change in varieties of European languages'.
Howell, R.B. 2006. 'Immigration and koineisation: the formation of early Modern Dutch urban vernaculars'. *Transactions of the Philological Society* 104(2): 207-227.
Karlsson, F., M. Miestamo & K. Sinnemäki (eds.), 'Introduction: The problem of language complexity'. In: M. Miestamo, K. Sinnemäki & F. Karlsson (eds.), *Language complexity. Typology, contact, change*. Amsterdam: Benjamins. vii-xiv.
Kusters, W. 2003. *Linguistic complexity: the influence of social change on verbal inflection*. Utrecht: LOT Dissertation Series.
Ladd, R.D., S.G. Roberts & D. Dediu. 2015. 'Correlation studies in typological and historical linguistics'. *Annual Review of Linguistics* 1: 221-241.
Lieberman, E., Michel, J.-B., J. Jackson, T. Tang, M.A. Nowak. 2007. 'Quantifying the evolutionary dynamics of language'. *Nature* 449: 713-716.
Lucas, C. 2015. 'Contact-induced change'. In: C. Bowern & B. Evans (eds.), *The Routledge handbook of historical linguistics*. London: Routledge. 519-536.
Lupyan, G. & R. Dale. 2010. 'Language structure is partly determined by social structure'. *PLoS ONE* 5(1).
Nettle, D. 1999. 'Is the rate of linguistic change constant?' *Lingua* 108: 119-136.
Pijpops, D, K. Beuls & F. Van de Velde. 2015. 'The rise of the verbal weak inflection in Germanic. An agent-based model'. *Computational Linguistics in the Netherlands Journal* 5: 81-102.
Roberts, S. & J. Winters. 2012. 'Social structure and language structure: the new nomothetic approach'. *Psychology of Language and Communication* 16(2): 89-112.
Steels, Luc. 2016. 'Agent-based models for the emergence and evolution of grammar'. *Phil. Trans. R. Soc. B*. 2015447.
Szmrecsanyi, B. 2012. Analyticity and Syntheticity in the history of English. In: T. Nevalainen, E. Traugott (eds.), *The Oxford handbook of the history of English*. Oxford: Oxford University Press. 654-665.
Thomason, S.G. 2003. 'Contact as a source of language change'. In: B.D. Joseph & R.D. Janda (eds.), The handbook of historical linguistics. Oxford: Blackwell. 686-712.
Trudgill, P. 2002. 'Linguistic and social typology'. In: J.K. Chambers, P. Trudgill & N. Schilling-Estes (eds.), *The handbook of language variation and language change*. Oxford: Blackwell. 707-728.
Weber, E. 1991. *My France, Politics, culture, myths*. Cambridge, MA: Harvard University Press.
Wexler, K. 1998. 'Very early parameter setting and the unique checking constraint: a new explanation of the optional infinitive stage'. *Lingua* 106:23-79
Wichmann, S. & E.W. Holman. 2009. 'Population Size and Rates of Language Change'. *Human Biology* 81: 259-274.
Woods, N. 2001. 'Internal and external dimensions of language change: the great divide? Evidence from New Zealand English'. *Linguistics* 39(5): 973-1007.